JOURNAL OF COMPLEXITY 17, 117–153 (2001)
doi:10.1006/jcom.2000.0570, available online at http://www.idealibrary.com on IDEAL®

The Optimal Discretization of Stochastic Differential Equations

Norbert Hofmann

Fachbereich Mathematik, Johann Wolfgang Goethe-Universität, Postfach 111932, 60054 Frankfurt, Germany
E-mail: hofmann@math.uni-frankfurt.de

and

Thomas Müller-Gronbach and Klaus Ritter

Fachbereich Mathematik, Technische Universität Darmstadt, Schloßgartenstraße 7,

Received April 20, 2000

We study pathwise approximation of scalar stochastic differential equations. The mean squared L_2-error and the expected number n of evaluations of the driving Brownian motion are used for the comparison of arbitrary methods. We introduce an adaptive discretization that reflects the local properties of every single trajectory. The corresponding error tends to zero like $c \cdot n^{-1/2}$, where c is the average of the diffusion coefficient in space and time. Our method is justified by the matching lower bound for arbitrary methods that are based on n evaluations on the average. Hence the adaptive discretization is asymptotically optimal. The new method is very easy to implement, and about 7 additional arithmetical operations are needed per evaluation of the Brownian motion. Hereby we can determine the complexity of pathwise approximation of stochastic differential equations. We illustrate the power of our method already for moderate accuracies by means of a simulation experiment.

Key Words: stochastic differential equation; pathwise approximation; adaptive discretization; asymptotic optimality; lower bounds; complexity.

0885-064X/01 $35.00

1. INTRODUCTION

Stochastic differential equations can be solved explicitly only in exceptional cases, and therefore numerical methods must be used in general. In this paper we study pathwise (or strong) approximation for stochastic differential equations, and we analyze numerical methods with respect to their error and computational cost.

Let W denote a one-dimensional Brownian motion on the unit interval, and consider a scalar stochastic differential equation

$$dX(t) = a(t, X(t))\, dt + \sigma(t, X(t))\, dW(t), \qquad t \in [0, 1], \tag{1}$$

with drift coefficient a, diffusion coefficient σ, and initial value $X(0)$. A pathwise approximation to the solution X is a stochastic process $\bar{X}$ whose paths are close to the respective paths of X.

Every numerical method relies on finite and therefore partial information about the underlying Brownian motion. We assume that W may be observed at adaptively chosen points

$$\tau_1, ..., \tau_\nu \in]0, 1].$$

The choice of τ_{k+1} may depend in any measurable way on the previously computed values $W(\tau_1), ..., W(\tau_k)$. In particular, we do not require $\tau_k < \tau_{k+1}$. The total number ν of observations may be determined by any measurable termination criterion. The discrete data about W may then be used in any measurable way to produce a pathwise approximation $\bar{X}$. An adaptive discretization should take into account

- (**D**) the drift and diffusion coefficient,
- (**T**) the particular trajectory,
- (**E**) the error criterion.

Here we use the following error criterion. For a given pair of corresponding trajectories of X and $\bar{X}$ we measure the distance in the L_2-norm, and we define the error $e(\bar{X})$ of the method $\bar{X}$ by averaging over all trajectories, i.e.,

$$e(\bar{X}) = (E(\|X - \bar{X}\|_2^2))^{1/2}. \tag{2}$$

As this definition depends on the discretization only via $\bar{X}$, it may serve as a basis for comparing different discretizations. Moreover, we characterize the quality of $\bar{X}$ globally on $[0, 1]$ and not only on a finite number of points.

As a rough measure for the computational cost of $\bar{X}$ we use the expected number $n(\bar{X})$ of observations of W. A more realistic measure also involves a count of the arithmetical operations performed by $\bar{X}$.

We address the following question: How much does it cost to achieve an error at most ε? A complete answer consists of two parts, an upper bound for a specific method and lower bounds for arbitrary methods. The minimal cost over all methods with $e(\bar{X}) \leqslant \varepsilon$ is called the ε-complexity of strong approximation of Eq. (1).

We introduce an adaptive discretization that reflects the local smoothness of the solution. The discretization is very easy to implement and has the convenient property that $\tau_k < \tau_{k+1}$ for all k. The smoothness of the solution at the point $(t, X(t))$ is determined by

$$E((X(t+\delta) - X(t))^2 \mid X(t)) = \sigma^2(t, X(t)) \cdot \delta + o(\delta); \tag{3}$$

hence $|\sigma|\,(t, X(t))$ might be called a conditional Hölder constant. It is reasonable to choose a small step-size if the conditional Hölder constant is large and vice versa. We basically take the step-size proportionally to the inverse of the current value of $|\sigma|$, and we use Euler and Milstein steps to move along the unit interval from left to right. Hereby we get a method $\hat{X}_h^{**}$ with input parameter $h > 0$ such that

$$\lim_{h \to 0} n(\hat{X}_h^{**})^{1/2} \cdot e(\hat{X}_h^{**}) = E\left(\int_0^1 |\sigma|(t, X(t))\, dt\right) \Big/ \sqrt{6};$$

see Theorem 1(i). In particular this is an asymptotic upper bound for the method $\hat{X}_h^{**}$.

The matching asymptotic lower bound reads

$$\liminf_{N \to \infty} N^{1/2} \cdot e(\bar{X}_N) \geqslant E\left(\int_0^1 |\sigma|(t, X(t))\, dt\right) \Big/ \sqrt{6}$$

for every sequence of methods $\bar{X}_N$ such that

$$n(\bar{X}_N) \leqslant N;$$

see Theorem 2(i). Our adaptive method $\hat{X}_h^{**}$ is therefore asymptotically optimal. The best order of convergence is $1/2$ in terms of $n(\bar{X})$, and the best asymptotic constant is given by the mean of the conditional Hölder constant in space and time. Note that the drift a and the initial value $X(0)$ are present implicitly in this constant.

The method $\hat{X}_h^{**}$ adjusts the number ν of evaluations of the given trajectory of W to the smoothness of the corresponding trajectory of X. Therefore ν is non-constant, in general, and $\hat{X}_h^{**}$ is a method of varying cardinality. No a priori bound for the computation time is available. We also study methods with fixed cardinality and methods with path-independent step-size control, and we present asymptotically optimal methods in both classes. It turns out that varying cardinality helps, and sometimes varying cardinality must be used to achieve a small error at reasonable average cost. Consider, for instance, the geometric Brownian motion. Here the asymptotic constant for $\hat{X}_h^{**}$ depends only linearly on the volatility, while the asymptotic constant for optimal methods with fixed cardinality depends exponentially on the volatility.

Our analysis shows that the local smoothness (3) is the key quantity. This fact was already known in much more generality for problems of approximation of stochastic processes; see, e.g., Ritter (2000) and the references therein. Let us stress two essential differences between both problems

(1) strong approximation for stochastic differential equations,

(2) approximation (reconstruction) of stochastic processes.

Discrete observations of the process X itself are used for the approximation of X in (2), while only W is observable in (1). The solution X depends non-linearly on W in (1), while there is a linear dependence (the identity) in (2). According to (3) one can compute the local smoothness rather accurately by a numerical method. This may also explain why optimal discretizations can be determined for stochastic differential equations, while such discretizations are unknown for ordinary differential equations.

Most of the results for strong (and weak) approximation for stochastic differential equations provide upper bounds for specific methods with unspecified constants. See, e.g., Bouleau and Lépingle (1994), Kloeden and Platen (1995), Milstein (1995), and Talay (1995) for results and numerous references. Only a few papers deal with asymptotic constants. We mention two results that deal with the mean squared error at the point $t = 1$ for autonomous equations.

For the first time, asymptotic constants were derived by Clark and Cameron (1980), who considered the case $\sigma = 1$. They studied a specific method, namely, the conditional expectation of $X(1)$ given the values of W at an equidistant discretization.

The first result concerning adaptivity with respect to (**D**) is due to Cambanis and Hu (1996), who considered discretizations that are given as quantiles of a fixed density on the unit interval. Cambanis and Hu characterize the asymptotically best choice of the density, which leads to the

smallest asymptotic constant. Moreover, they pose several open problems, some of which are settled by the results of this paper. In particular, Theorem 2(ii) and (iii) solves problem (1.18) from Cambanis and Hu (1996).[1]

The first result concerning adaptivity with respect to (**T**) is due to Newton (1990), who determined the discretization from passage times of the driving process. This contrasts our approach, since we aim at trajectorial properties of the solution.

In Hofmann *et al.* (2000a, 2000b) we study pathwise approximation for equations (1) with additive noise, i.e., $\sigma(t, x) = \sigma(t)$. In this case it suffices to use adaption with respect to (**D**) and (**E**).

We also mention Gaines and Lyons (1997) and Mauthner (1998), who constructed an adaptive method and proved its convergence.

Another problem, which also arises in stochastic analysis, is the approximate computation of stochastic integrals. Wasilkowski and Woźniakowski (2000) and Hertling (2000) determine the complexity of this problem.

The paper is organized as follows. In Section 2 we define several versions of our adaptive method, namely, with varying or fixed cardinality and with a path-independent step-size control. In Section 3 we specify our assumptions regarding the equation. The drift and diffusion coefficients must satisfy Lipschitz conditions, and the initial value must have a finite fourth moment. Section 4 contains the asymptotic results for our methods. Moreover, we present the asymptotic analysis for the Milstein scheme based on an equidistant discretization. In Section 5 we formally define arbitrary methods for strong approximation and we present the lower bounds for all methods with varying cardinality, fixed cardinality, and path-independent step-size control. Moreover, we state the lower bound for methods based on an equidistant discretization. In Section 6 we combine the results from Sections 4 and 5 to determine the ε-complexity. The gap in the upper and lower bounds is only a factor of at most 8. This is due to the fact that asymptotically the method $\hat{X}_n^{**}$ only needs 7 additional arithmetical operations per evaluation of W. Hence our bounds give a very precise estimate of the actual computation time. Section 7 is devoted to a simulation experiment for the geometric Brownian motion $dX(t) = 3 \cdot X(t)\, dW(t)$ with $X(0) = 1$. According to the analysis, our adaptive method reduces the computation time by a factor 900, compared to an equidistant discretization. This is confirmed by the experiments already for moderate accuracies $\varepsilon \leqslant 2 \cdot 10^{-2}$. Proofs are given in Section 8, and the Appendix contains an upper bound for the error of the Milstein method, which is useful for our analysis.

[1] The answer is $\gamma = 1$ as well as $F(a, b) = C^2/6$ if only adaption with respect to (**D**) is used and $F(a, b) = (C^*)^2/6$ in general. Discretizations of varying size are not mentioned by Cambanis and Hu (1996).

2. THE ADAPTIVE METHOD

The smoothness of the solution X of (1) at the point $(t, X(t))$ is determined by the conditional Hölder constant $|\sigma|(t, X(t))$, see (3). Therefore it is reasonable to decrease the step-size $\tau_{\ell+1} - \tau_\ell$ with increasing value of $|\sigma|(\tau_\ell, X(\tau_\ell))$. Hereby we deal with the first two requirements (**D**) and (**T**) for an adaptive discretization. The precise relation between the step-size and the conditional Hölder constant depends on the error criterion. For the L_2-error (2) we basically take steps proportionally to $1/|\sigma|(\tau_\ell, X(\tau_\ell))$. Since $X(\tau_\ell)$ is unknown in general, an approximation must be used.

We take the Milstein scheme to compute approximations at discrete points. For every discretization

$$0 = \tau_0 < \cdots < \tau_m = 1 \tag{4}$$

this scheme is defined by

$$\hat{X}(\tau_0) = X(0)$$

and

$$\begin{aligned}\hat{X}(\tau_{\ell+1}) = \hat{X}(\tau_\ell) &+ a(\tau_\ell, \hat{X}(\tau_\ell)) \cdot (\tau_{\ell+1} - \tau_\ell) \\ &+ \sigma(\tau_\ell, \hat{X}(\tau_\ell)) \cdot (W(\tau_{\ell+1}) - W(\tau_\ell)) \\ &+ 1/2 \cdot (\sigma \cdot \sigma^{(0,1)})(\tau_\ell, \hat{X}(\tau_\ell)) \cdot ((W(\tau_{\ell+1}) - W(\tau_\ell))^2 - (\tau_{\ell+1} - \tau_\ell)),\end{aligned}$$

where $\ell = 0, ..., m-1$. Here $\sigma^{(0,1)}$ denotes the partial derivative of σ with respect to the second or state variable. A global approximation $\hat{X}$ for X on $[0, 1]$ is obtained by piecewise linear interpolation of the data $(\tau_\ell, \hat{X}(\tau_\ell))$. Obviously $\hat{X}$ depends on W only through its values at the discretization points.

We present several versions of our adaptive method.

2.1. *The Fully Adaptive Method $\hat{X}_h$.* We choose a basic step-size

$$h > 0$$

and define the adaptive step-size control by $\tau_0 = 0$ and

$$\tau_{\ell+1} = \tau_\ell + h/|\sigma|(\tau_\ell, \hat{X}(\tau_\ell))$$

as long as the right-hand side does not exceed one. Otherwise we put $\tau_{\ell+1}=1$. We refer to the corresponding Milstein approximation by $\hat{X}_h$ in the sequel. See Remark 3 for further discussion.

Obviously the adaptive method $\hat{X}_h$ is very easy to implement. Unfortunately, it seems to be difficult to analyze this method. Therefore we switch to a closely related method $\hat{X}_h^{**}$. Simulation experiments indicate that the fully adaptive method is superior to $\hat{X}_h^{**}$ for moderate basic step sizes. Our analysis shows that $\hat{X}_h$ cannot be better than $\hat{X}_h^{**}$ asymptotically. In fact, we conjecture that both methods have the same asymptotic performance.

2.2. *The Simplified Adaptive Method* $\hat{X}_h^{**}$. For $h>0$ take $k_h \in \mathbb{N}$ such that

$$\lim_{h\to 0} k_h \cdot h = 0 \tag{5}$$

and

$$\lim_{h\to 0} k_h^2 \cdot h = \infty. \tag{6}$$

The modification $\hat{X}_h^{**}$ of $\hat{X}_h$ does not update the step-size in every step but only at the equidistant points

$$\tau_{i,0} = i/k_h, \qquad i=0, ..., k_h-1.$$

Suppose that

$$x_i = \hat{X}_h^{**}(\tau_{i,0})$$

is already computed, where $x_i = X(0)$ for $i=0$. Put

$$\sigma_i = \sigma(\tau_{i,0}, x_i), \qquad a_i = a(\tau_{i,0}, x_i),$$

and define an adaptive discretization of the subinterval $]\tau_{i,0}, \tau_{i+1,0}[$ by

$$\tau_{i,j+1} = \tau_{i,j} + h/|\sigma_i|$$

as long as the right-hand side is less than $\tau_{i+1,0}$. Here $\tau_{k_h,0}=1$. On the subinterval $]\tau_{i,0}, \tau_{i+1,0}[$ we use the Euler method

$$\hat{X}_h^{**}(\tau_{i,j+1}) = \hat{X}_h^{**}(\tau_{i,j}) + a_i \cdot (\tau_{i,j+1} - \tau_{i,j}) + \sigma_i \cdot (W(\tau_{i,j+1}) - W(\tau_{i,j})),$$

without updating the drift and diffusion coefficient. The approximate solution $x_{i+1} = \hat{X}_h^{**}(\tau_{i+1,0})$ at the right endpoint is computed by a single Milstein step of length $1/k_h$ starting at the left endpoint $\tau_{i,0}$ with initial value x_i. Of course, the Milstein step is based on $W(\tau_{i+1,0}) - W(\tau_{i,0})$, which is the sum of the Brownian increments $W(\tau_{i,j+1}) - W(\tau_{i,j})$. Globally we use piecewise linear interpolation. In Theorem 1(i) we present the asymptotic analysis of the error $e(\hat{X}_h^{**})$ with h tending to zero.

Note that on each subinterval $]\tau_{i,0}, \tau_{i+1,0}[$ we apply the Euler method to the stochastic differential equation

$$dX(t) = a_i\, dt + \sigma_i\, dW_i(t)$$

with initial value $X(\tau_{i,0}) = x_i$ and driving Brownian motion $W_i(t) = W(t) - W(\tau_{i,0})$. For this equation the Euler and the Milstein method coincide, and they yield the exact solution at the discretization points.

It is known that the Milstein steps of length $1/k_h$ yield a squared error of order $1/k_h^2$ at all points $\tau_{i,0}$. In contrast, Euler steps of length $h/|\sigma_i|$ only yield a squared error of order h at $\tau_{i+1,0}$ in general, even if $X(\tau_{i,0})$ were known. The use of the large Milstein steps is therefore crucial for the good approximation of the diffusion coefficient at $\tau_{i+1,0}$.

The total number of observations of W that are used by $\hat{X}_h^{**}$ depends on the particular trajectory, such that $\hat{X}_h^{**}$ is a method of varying cardinality. The number of observations is roughly given by $1/h \cdot S$, where

$$S = \int_0^1 |\sigma|\,(t, X(t))\, dt \tag{7}$$

is the average of the conditional Hölder constant for the particular trajectory. The quantity S may heavily depend on the trajectory, see Fig. 2 in Section 7, and consequently, there is no a priori bound on the computation time available for the user. If all approximations have to be computed in the same amount of time, we suggest to use the following version $\hat{X}_n^*$ of the adaptive method. We stress, however, that one has to pay a price for this property, see Theorems 1(i) and 2(ii) as well as Example 1.

2.3. *The Adaptive Method $\hat{X}_n^*$ with Fixed Cardinality n.* For $n \in \mathbb{N}$ take $k_n \in \mathbb{N}$ such that

$$\lim_{n \to \infty} k_n/n = 0 \tag{8}$$

and

$$\lim_{n\to\infty} k_n^2/n = \infty. \tag{9}$$

In contrast to $\hat{X}_h^{**}$ and $\hat{X}_h$, the method $\hat{X}_n^*$ does not progress from left to right. At first the Milstein approximation

$$x_i = \hat{X}_n^*(\tau_{i,0})$$

is computed at all equidistant points

$$\tau_{i,0} = i/k_n, \qquad i = 0, ..., k_n - 1.$$

Thereafter the discretization is refined adaptively in the following way. Put

$$\sigma_i = \sigma(\tau_{i,0}, x_i)$$

and let

$$\mu_i = \begin{cases} \left\lfloor (n-k_n)\cdot|\sigma_i| \Big/ \sum_{\ell=0}^{k_n-1} |\sigma_\ell| \right\rfloor, & \text{if } \sum_{\ell=0}^{k_n-1} |\sigma_\ell| > 0 \\ \lfloor (n-k_n)/k_n \rfloor, & \text{otherwise.} \end{cases}$$

The adaptive discretization of the subinterval $]\tau_{i,0}, \tau_{i+1,0}[$ is given by

$$\tau_{i,j+1} = \tau_{i,j} + 1/(k_n \cdot (\mu_i + 1))$$

as long as the right-hand side is less than $\tau_{i+1,0}$. Here $\tau_{k_n,0} = 1$. Now $\hat{X}_n^*$ proceeds in the same way as $\hat{X}_h^{**}$. The asymptotic behavior of the error $e(\hat{X}_n^*)$ is determined in Theorem 1(ii).

Note that the number of Euler steps on $]\tau_{i,0}, \tau_{i+1,0}[$ is given by μ_i. Since

$$n - 2k_n \leqslant k_n - 1 + \sum_{i=0}^{k_n-1} \mu_i \leqslant n,$$

the method $\hat{X}_n^*$ uses at most n observations of W for every trajectory. Due to (8), the upper bound n is sharp, asymptotically. By formally introducing a few additional observations we get a method of fixed cardinality n. Observe that $\hat{X}_n^*$ takes steps of size roughly given by $1/n \cdot S/|\sigma|\,(\tau_{i,0}, X(\tau_{i,0}))$ on $]\tau_{i,0}, \tau_{i+1,0}[$.

2.4. *The Adaptive Method $\hat{X}_n$ with Path-Independent Step-Size Control.* Now we introduce a method $\hat{X}_n$ that is adaptive only with respect to the equation. The method $\hat{X}_n$ coincides with $\hat{X}_n^*$ up to the fact that σ_i is replaced by

$$(E(\sigma^2(\tau_{i,0}, X(\tau_{i,0}))))^{1/2}$$

for $i=0, ..., k_n-1$ in the definition of the numbers μ_i. In Theorem 1(iii) we present the asymptotic analysis of the error $e(\hat{X}_n)$.

In contrast to the methods from Subsections 2.1–2.3, $\hat{X}_n$ is not easy to implement, as it requires knowledge of $E(\sigma^2(t, X(t)))$ at the points $t=\tau_{i,0}$.

3. ASSUMPTIONS

Throughout this paper we assume that the drift and diffusion coefficients

$$a, \sigma: [0, 1] \times \mathbb{R} \to \mathbb{R}$$

and the initial value $X(0)$ have the following properties.

(A) Both, a and σ are differentiable with respect to the state variable. Moreover, there exists a constant $K>0$ such that $f=a$ and $f=\sigma$ satisfy

$$\begin{aligned} |f(t, x)-f(t, y)| &\leqslant K \cdot |x-y|, \\ |f(s, x)-f(t, x)| &\leqslant K \cdot (1+|x|) \cdot |s-t|, \\ |f^{(0,1)}(t, x)-f^{(0,1)}(t, y)| &\leqslant K \cdot |x-y| \end{aligned}$$

for all $s, t \in [0, 1]$ and $x, y \in \mathbb{R}$.

(B) The initial value $X(0)$ is independent of W and

$$E(X(0))^4 < \infty.$$

Note that (A) yields the linear growth condition

$$|f(t, x)| \leqslant c \cdot (1+|x|).$$

Moreover, $f^{(0,1)}$ is bounded and

$$|f(t, x)-f(t, y)-f^{(0,1)}(t, y)(x-y)| \leqslant c \cdot (x-y)^2.$$

Given the above properties, a pathwise unique strong solution of Eq. (1) with initial value $X(0)$ exists. In particular the conditions assure that

$$\sup_{t \in [0,1]} E(X(t))^4 < \infty. \tag{10}$$

4. ANALYSIS OF THE ADAPTIVE METHOD

To every Eq. (1) we associate the constants

$$C^{**} = E\left(\int_0^1 |\sigma|(t, X(t))\, dt\right),$$

$$C^{*} = \left(E\left(\int_0^1 |\sigma|(t, X(t))\, dt\right)^2\right)^{1/2},$$

$$C = \int_0^1 (E(\sigma^2(t, X(t))))^{1/2}\, dt.$$

Let $n(\hat{X}_h^{**})$ denote the average cardinality of $\hat{X}_h^{**}$, i.e., the expected number of observations of W that are used by this method. We relate the average cardinality to the error $e(\hat{X}_h^{**})$ as h tends to zero. Recall that $\hat{X}_n^*$ and $\hat{X}_n$ are of fixed cardinality n.

THEOREM 1. *The adaptive methods* $\hat{X}_h^{**}$, $\hat{X}_n^*$, *and* $\hat{X}_n$ *satisfy*

(i) $\lim_{h \to 0} n(\hat{X}_h^{**})^{1/2} \cdot e(\hat{X}_h^{**}) = C^{**}/\sqrt{6}$,

(ii) $\lim_{n \to \infty} n^{1/2} \cdot e(\hat{X}_n^{*}) = C^{*}/\sqrt{6}$,

(iii) $\lim_{n \to \infty} n^{1/2} \cdot e(\hat{X}_n) = C/\sqrt{6}$

for every Eq. (1).

Clearly, one can hardly justify the use of an adaptive method, if it is not superior to good methods that are based on an equidistant discretization. At first we take the equidistant Milstein scheme $\hat{X}_n^{\text{equi}}$ with step-size $1/n$ for the comparison. In the general approach in Section 5 we establish the optimality of the methods $\hat{X}_h^{**}$, $\hat{X}_n^*$, and $\hat{X}_n$, and we show in particular

their superiority to every method that is based on an equidistant discretization. We define

$$C^{\text{equi}} = \left(\int_0^1 E(\sigma^2(t, X(t)))\, dt \right)^{1/2}$$

for the analysis of the method $\hat{X}_n^{\text{equi}}$.

PROPOSITION 1. *The equidistant Milstein scheme* $\hat{X}_n^{\text{equi}}$ *satisfies*

$$\lim_{n \to \infty} n^{1/2} \cdot e(\hat{X}_n^{\text{equi}}) = C^{\text{equi}}/\sqrt{6}$$

for every Eq. (1).

Note that the order of convergence is 1/2 for all the above methods and

$$C^{**} \leqslant C^{*} \leqslant C \leqslant C^{\text{equi}}$$

with strict inequality in most cases. In fact, huge differences may occur between these constants, as shown in Example 1. See Remark 2 for a characterization of equality.

EXAMPLE 1. The asymptotic constants from Theorem 1 and Proposition 1 can be determined analytically only in exceptional cases, e.g., for linear equations or equations with additive noise. Here we consider the linear equation

$$dX(t) = bX(t)\, dW(t)$$

with initial value $X(0) = 1$. The solution is the geometric Brownian motion

$$X(t) = \exp(-b^2/2 \cdot t + b \cdot W(t))$$

with drift zero. Straightforward calculations yield

$$\begin{aligned} C^{**} &= b, \\ C^{*} &= (2 \cdot (\exp(b^2) - b^2 - 1))^{1/2}/b, \\ C &= 2/b \cdot (\exp(b^2/2) - 1), \\ C^{\text{equi}} &= (\exp(b^2) - 1)^{1/2}. \end{aligned}$$

For $\hat{X}_n^{\text{equi}}$, $\hat{X}_n$, and $\hat{X}_n^{*}$ the asymptotic constant depends exponentially on the volatility b. For $\hat{X}_h^{**}$ we only have a linear dependence on b.

5. LOWER BOUNDS AND OPTIMALITY OF THE ADAPTIVE METHOD

Theorem 1 and Proposition 1 determine the asymptotic performance of specific methods for arbitrary equations. These methods are based on a realization of the initial value $X(0)$ and on a finite number of observations of a trajectory of the driving Brownian motion W. The choice of the adaptive schemes will now be justified by lower bounds that hold for arbitrary methods of the above form and arbitrary equations.

Fix a and σ, and consider the corresponding Eq. (1). Formally a general method is then defined by mappings

$$\psi_k: \mathbb{R}^k \to [0, 1],$$
$$\chi_k: \mathbb{R}^{k+1} \to \{\text{STOP}, \text{GO}\},$$
$$\phi_k: \mathbb{R}^{k+1} \to L_2([0, 1])$$

for $k \in \mathbb{N}$. The sequential observation of a trajectory w starts at the knot $\psi_1(x)$, which may depend on the realization x of the initial value. After k steps we have observed the data

$$\Psi_k(x, w) = (x, y_1, ..., y_k),$$

where

$$y_1 = w(\psi_1(x)), ..., y_k = w(\psi_k(x, y_1, ..., y_{k-1})).$$

A decision to stop or to further evaluate w is made after each step, and the total number of observations of w is given by

$$\nu(x, w) = \min\{k \in \mathbb{N} : \chi_k(\Psi_k(x, w)) = \text{STOP}\}.$$

If $\nu(x, w) < \infty$ then the data

$$\Psi(x, w) = \Psi_{\nu(x, w)}(x, w)$$

are used to construct the approximation $\phi_{\nu(x, w)}(\Psi(x, w))$.

We only assume measurability of the mappings ψ_k, χ_k, and ϕ_k. For simplicity, these mappings are defined on the whole spaces $\mathbb{R}^k$ and $\mathbb{R}^{k+1}$,

respectively. Obviously only the case $v(X(0), W) < \infty$ with probability one is of practical interest. Then we end up with the method

$$\bar{X} = \phi_{v(X(0),\, W)}(\Psi(X(0), W)).$$

As previously we relate the error $e(\bar{X})$ to the average cardinality

$$n(\bar{X}) = E(v(X(0), W)).$$

Let $\mathbb{X}^{**}$ denote the class of all methods of the above form, and put

$$\mathbb{X}_N^{**} = \{\bar{X} \in \mathbb{X}^{**} : n(\bar{X}) \leqslant N\}$$

for $N \in \mathbb{N}$. Clearly $\hat{X}_h$, $\hat{X}_h^{**} \in \mathbb{X}^{**}$ for the adaptive methods from Subsection 2.1 and 2.2. The quantity

$$e^{**}(N) = \inf\{e(\bar{X}) \colon \bar{X} \in \mathbb{X}_N^{**}\}$$

is the minimal error that can be obtained by methods that use at most N sequential observations of W on the average.

As a subclass $\mathbb{X}^* \subset \mathbb{X}^{**}$ we consider all methods that use the same number of observations for all trajectories. Formally this means that the mappings χ_k are constant and the cardinality $v = \min\{k \in \mathbb{N} : \chi_k = \text{STOP}\}$ is fixed. We put

$$\mathbb{X}_N^* = \{\bar{X} \in \mathbb{X}^* : n(\bar{X}) \leqslant N\}$$

as well as

$$e^*(N) = \inf\{e(\bar{X}) : \bar{X} \in \mathbb{X}_N^*\}.$$

Recall that the adaptive method $\hat{X}_n^*$ from Subsection 2.3 uses n observations for each trajectory, and therefore $\hat{X}_n^* \in \mathbb{X}_n^*$.

The subclass $\mathbb{X} \subset \mathbb{X}^*$ consists of all methods that use the same observation points for every trajectory. Formally the mappings ψ_k and χ_k are constant, such that $\Psi(x, w) = (x, w(\psi_1), ..., w(\psi_v))$. We put

$$\mathbb{X}_N = \{\bar{X} \in \mathbb{X} : \ n(\bar{X}) \leqslant N\}$$

as well as

$$e(N) = \inf\{e(\bar{X}) : \ \bar{X} \in \mathbb{X}_N\}.$$

Note that the method $\hat{X}_n$ from Section 2.4 belongs to the class $\mathbb{X}_n$.

THEOREM 2. *The minimal errors satisfy*

(i) $\lim_{N\to\infty} N^{1/2} \cdot e^{**}(N) = C^{**}/\sqrt{6}$,
(ii) $\lim_{N\to\infty} N^{1/2} \cdot e^{*}(N) = C^{*}/\sqrt{6}$,
(iii) $\lim_{N\to\infty} N^{1/2} \cdot e(N) = C/\sqrt{6}$

for every Eq. (1).

From Theorems 1 and 2 we immediately get the following.

Main Result. Suppose that $C^{**} > 0$. Then the methods $\hat{X}_h^{**}$, $\hat{X}_n^{*}$ and $\hat{X}_n$ are asymptotically optimal in the respective classes $\mathbb{X}_N^{**}$ with $N = n(\hat{X}_h^{**})$, $\mathbb{X}_n^{*}$ and $\mathbb{X}_n$.

Finally, we consider arbitrary methods that are based on equidistant discretizations, and we define

$$e^{\text{equi}}(N) = \inf\{e(\phi(X(0), W(1/N), ..., W(1))): \phi: \mathbb{R}^{n+1} \to L_2([0, 1]) \text{ measurable}\}.$$

It is known that the conditional expectation yields the optimal ϕ. We get a lower bound for e^{equi} from Theorem 2(iii) and an upper bound from Proposition 1. The upper bound turns out to be sharp.

PROPOSITION 2. *For every Eq.* (1),

$$\lim_{N\to\infty} N^{1/2} \cdot e^{\text{equi}}(N) = C^{\text{equi}}/\sqrt{6}.$$

Therefore the equidistant Milstein scheme is asymptotically optimal among all methods that are based on equidistant discretizations.

Remark 1. The conditional Hölder constant $|\sigma|(t, X(t))$ describes the smoothness of X locally in time and space. The constants C^{**} and C^{*} are based on the average S of the conditional Hölder constant along a trajectory, see (7). We have

$$C^{**} = E(S), \qquad C^{*} = (E(S^2))^{1/2}.$$

Due to (3),

$$E(X(t+\delta)-X(t))^2=\alpha^2(t)\cdot\delta+o(\delta)$$

with

$$\alpha(t)=(E(\sigma^2(t, X(t))))^{1/2}. \tag{11}$$

Hence $\alpha(t)$ describes the smoothness of X only locally in time. The constants C and C^{equi} are based on α. We have

$$C=\int_0^1 \alpha(t)\,dt, \qquad C^{\text{equi}}=\left(\int_0^1 \alpha^2(t)\,dt\right)^{1/2}.$$

Remark 2. Clearly $C=C^{\text{equi}}$ iff α is constant. Furthermore, $C^*=C$ iff there exist $t_0\in[0,1]$ and $\gamma\in C([0,1])$ such that, with probability one,

$$\forall t\in[0,1], \qquad |\sigma|\,(t, X(t))=\gamma(t)\cdot|\sigma|\,(t_0, X(t_0)). \tag{12}$$

Note that γ is then determined by

$$\gamma=\alpha/(E(\sigma^2(t_0, X(t_0))))^{1/2}$$

if $C>0$.

Finally, the Markov property of X implies that $C^{**}=C^*$ iff, with probability one,

$$\forall t\in[0,1], \qquad |\sigma|(t, X(t))=\alpha(t). \tag{13}$$

Obviously (13) implies (12), and both conditions are equivalent in the case of a nonzero and deterministic $\sigma(0, X(0))$. It would be nice to know whether this equivalence holds for general Eq. (1).

Note that (13) is satisfied for equations with additive noise, i.e., $\sigma^{(0,1)}=0$ implies $C^{**}=C$. Moreover, the Milstein scheme and the Euler scheme coincide in this situation. Equations with additive noise were studied in Hofmann *et al.* (2000a).

Finally, the constants C^{**}, C^*, C, and C^{equi} vanish altogether iff $C^{**}=0$, which is equivalent to $\sigma=0$ with probability one. Thus we have excluded ordinary differential equations in the Main Result.

Remark 3. The error criterion does influence the optimal step-size control. If small errors

$$e(\bar{X}) = (E\,\|X - \bar{X}\|_\infty^q)^{1/q}, \qquad 1 \leqslant q < \infty,$$

with respect to the L_∞-norm on $[0, 1]$ are needed, then steps of length $h/\sigma^2(\tau_\ell, \hat{X}(\tau_\ell))$ seem to be appropriate. At least for (systems of) stochastic differential equations with additive noise, this choice leads to asymptotically optimal methods. See Hofmann *et al.* (2000b).

For practical purposes, the definition of the step-size control in Subsection 2.1 must be extended, as $|\sigma|(\tau_\ell, \hat{X}(\tau_\ell))$ might be too small or even zero. A simple modification is given by

$$\tau_{\ell+1} = \tau_\ell + \min(h/|\sigma|(\tau_\ell, \hat{X}(\tau_\ell), h^{2/3})$$

for errors in L_2-norm. See Hofmann *et al.* (2000a) for a motivation of this choice in case of equations with additive noise. For errors in L_∞-norm $h^{2/3}$ may be replaced by $h \cdot \ln h^{-1}$, see Hofmann *et al.* (2000b).

6. THE COMPLEXITY OF STOCHASTIC DIFFERENTIAL EQUATIONS

The ε-complexity $\mathrm{comp}(\varepsilon)$ of a numerical problem is the minimal computational cost to solve the problem with error at most ε. See Traub *et al.* (1988). In this paper the problem is pathwise approximation for a stochastic differential equation, and the error is the mean squared L_2-distance. For a particular method and trajectory the cost is determined by the following quantities

(1) the number of evaluations of W,

(2) the number of evaluations of the drift and diffusion coefficients,

(3) the computational cost to evaluate the mappings ψ_k, χ_k, and ϕ_k.

We assume that W can be evaluated at cost c_1 at any point from $[0, 1]$, and that the drift and diffusion coefficients can be evaluated at cost c_2 at any point from $[0, 1] \times \mathbb{R}$. Moreover, we assume that arithmetical and similar operations, which are needed to evaluate ψ_k, χ_k, and ϕ_k, are carried out at unit cost. It is reasonable to assume $\min\{c_1, c_2\} > 1$. For some equations, c_2 should be much larger than one. For every method $\bar{X}$ we let $c(\bar{X})$ denote the expected cost over all trajectories.

Clearly

$$c_1 \cdot n(\bar{X}) \leqslant c(\bar{X}),$$

but often $c(\bar{X})$ is much larger than $c_1 \cdot n(\bar{X})$, for instance, if ordinary differential equations or optimization problems are solved in (3) or weak approximations for stochastic differential equations are computed. See Subsection 2.4 for a method of the latter kind.

Now we look at the adaptive method $\hat{X}_h^{**}$. By Lemma 7, $n(\hat{X}_h^{**})$ is of order $1/h$, and $k_h = o(1/h)$ by (5). Hence the cost from (2) is irrelevant asymptotically. We need one arithmetical operation to obtain the Brownian increment and one more operation to accumulate the Brownian increment for the Milstein step. Moreover, three arithmetical operations are needed per Euler step. Finally the discretization point is updated and a comparison is made. We thus conclude that

$$c_1 \leqslant \liminf_{h \to \infty} \frac{c(\hat{X}_h^{**})}{n(\hat{X}_h^{**})} \leqslant \limsup_{h \to \infty} \frac{c(\hat{X}_h^{**})}{n(\hat{X}_h^{**})} \leqslant c_1 + 7.$$

Hence the cost of $\hat{X}_h^{**}$ is determined by its cost from (1), up to a multiplicative constant that is close to one. Together with Theorems 1 and 2 this yields the following result.

THEOREM 3. *The ε-complexity of every Eq.* (1) *satisfies*

$$\liminf_{\varepsilon \to 0} \varepsilon^2 \cdot \mathrm{comp}(\varepsilon) \geqslant \frac{c_1}{6} \cdot (C^{**})^2$$

and

$$\limsup_{\varepsilon \to 0} \varepsilon^2 \cdot \mathrm{comp}(\varepsilon) \leqslant \frac{c_1 + 7}{6} \cdot (C^{**})^2.$$

Remark 4. Although the methods $\hat{X}_h^{**}$ and $\hat{X}_n^{*}$ are adaptive, they are very well suited for parallel processing. In fact, only the evaluation of W at the points $\tau_{i,0}$ must be done sequentially. Thereafter the Euler steps can be performed in parallel on all subintervals.

7. SIMULATION EXPERIMENTS

So far we have presented an asymptotic analysis where n, the (expected) number of evaluations of W, tends to infinity. Let us now illustrate the

practical relevance of these results for moderate size of n or ε, respectively, by means of a simulation experiment. We take the geometric Brownian motion from Example 1 with

$$b = 3.$$

Note that the asymptotic constants are known in this case.

For the fully adaptive method $\hat{X}_h$ the error e as well as n are shown in Table I for several values of h. Both quantities are determined by simulations (simul.). We simulate a sufficiently large number K of trajectories of the driving Brownian motion, choose a sufficiently accurate piecewise linear interpolation $\tilde{X}$ of the geometric Brownian motion X and estimate e according to

$$e = \left(\frac{1}{K} \cdot \sum_{i=1}^{K} \| \tilde{X}(w_i) - \hat{X}_h(w_i) \|_2^2 \right)^{1/2}.$$

Furthermore we use the mean number of knots for n.

For comparison we take the simplified adaptive method $\hat{X}_h^{**}$, where we use the asymptotic formulas (asymp.) for e and n, see Theorem 1(i) and Lemma 7. Table I supports our conjecture that Theorem 1(i) also holds for $\hat{X}_h$ instead of $\hat{X}_h^{**}$.

For several methods or classes of methods we show how n depends on the error, which varies within the reasonable range $[1 \cdot 10^{-3}, 2 \cdot 10^{-2}]$. See Fig. 1. Solid or dotted lines are computed by means of asymptotic formulas, and ★ corresponds to the simulated values from Table I. Finally + and ○ are computed by use of explicit formulas for $e(\hat{X}_h^{**})$ and $n(\hat{X}_h^{**})$ as well as $e(\hat{X}_n^{\text{equi}})$. Such explicit formulas can be derived in the particular case of the geometric Brownian motion.

TABLE I

Error and Expected Number of Evaluations for $\hat{X}_h$ and $\hat{X}_h^{**}$

h	e		n	
	Simul.	Asymp.	Simul.	Asymp.
$5 \cdot 10^{-4}$	$1.82 \cdot 10^{-2}$	$1.58 \cdot 10^{-2}$	6 009	6 000
$4 \cdot 10^{-4}$	$1.59 \cdot 10^{-2}$	$1.41 \cdot 10^{-2}$	7 624	7 500
$2 \cdot 10^{-4}$	$1.04 \cdot 10^{-2}$	$1.00 \cdot 10^{-2}$	14 981	15 000
$1 \cdot 10^{-4}$	$7.09 \cdot 10^{-3}$	$7.07 \cdot 10^{-3}$	29 843	30 000
$6 \cdot 10^{-5}$	$5.55 \cdot 10^{-3}$	$5.48 \cdot 10^{-3}$	51 297	50 000
$3 \cdot 10^{-5}$	$3.81 \cdot 10^{-3}$	$3.87 \cdot 10^{-3}$	100 122	100 000

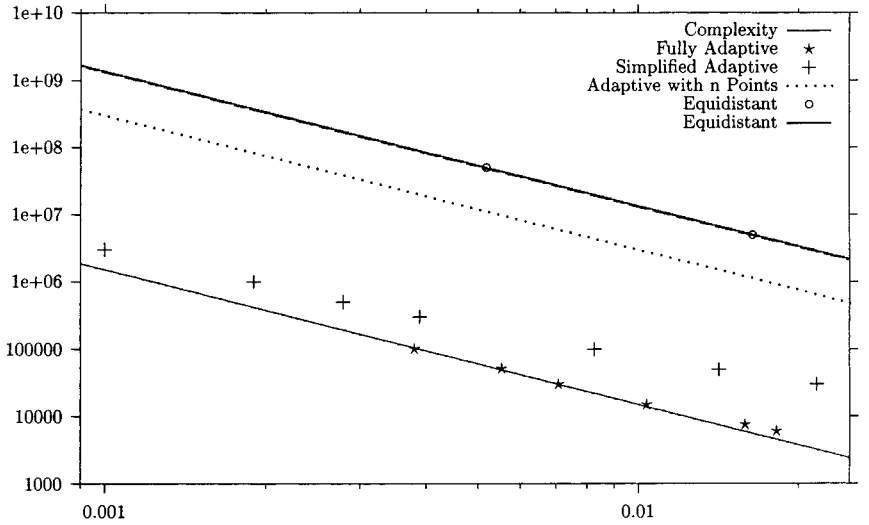

FIG. 1. Expected number of evaluations vs error.

The fully adaptive method gets very close to the asymptotic lower bound from Theorem 3 (complexity). The simplified adaptive method is not quite as good as the fully adaptive one. A significant loss of performance occurs if we require that n is fixed for all trajectories; the asymptotic formula from Theorem 2(ii) is used here. For equidistant discretizations the asymptotic formula from Proposition 2 yields a very precise approximation for the error $e(\hat{X}_n^{\text{equi}})$.

We conclude in particular that the computation time increases by a factor close to

$$(\exp(9)-1)/9 = 900.2...,$$

if we wish to achieve the same error by an equidistant discretization instead of a fully adaptive one. The absolute values of n demonstrate the practical relevance of this fact.

Finally we consider again the fully adaptive method $\hat{X}_h$. We show that the number $v(1, w)$ of evaluations of the trajectory w strongly depends on w. To this end relative frequencies are presented in Fig. 2 in the case $h = 2 \cdot 10^{-4}$. Recall that the mean $n(\hat{X}_h^{**})$ of v is about 15 000; see Table I. The median of v is about 4 450. Two percent of the trajectories required less than 1000 evaluations. Less than 0.1% of the trajectories required more than 10^6 evaluations. The maximal number $v(1, w)$ that occurred in the simulation was about $1.41 \cdot 10^7$, and the empirical deviation of $v(1, w)$ was $9.2 \cdot 10^4$.

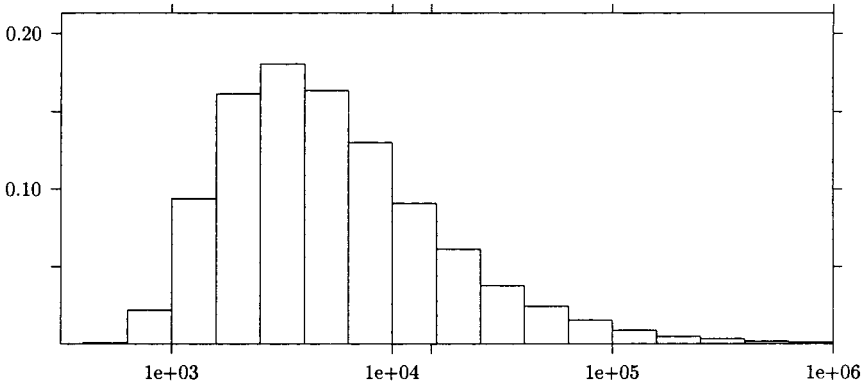

FIG. 2. Relative frequencies of v for $\hat{X}_h$ with $h = 2 \cdot 10^{-4}$.

8. PROOFS

Instead of estimating $X - \bar{X}$ directly, we introduce processes $\check{X}_m$ and consider $X - \check{X}_m$ as well as $\check{X}_m - \bar{X}$ separately. For $m \in \mathbb{N}$ let

$$t_\ell = \ell/m, \qquad \ell = 0, ..., m.$$

The process $\check{X}_m$ is given by $\check{X}_m(0) = X(0)$ and

$$\check{X}_m(t) = \check{X}_m(t_\ell) + a(t_\ell, \check{X}_m(t_\ell)) \cdot (t - t_\ell) + \sigma(t_\ell, \check{X}_m(t_\ell)) \cdot (W(t) - W(t_\ell)) + 1/2 \cdot (\sigma \cdot \sigma^{(0,1)})(t_\ell, \check{X}_m(t_\ell)) \cdot ((W(t) - W(t_\ell))^2 - (t - t_\ell)) \tag{14}$$

for $t \in [t_\ell, t_{\ell+1}]$. Note that $\check{X}_m$ coincides with the Milstein scheme at the discretization points t_ℓ. Complete knowledge of a trajectory of W is needed to generate the corresponding trajectory of $\check{X}_m$. Therefore $\check{X}_m$ is not an implementable numerical scheme for the global approximation of X. By Theorem 4, $E(X(t) - \check{X}_m(t))^2 = O(m^{-2})$ holds uniformly with respect to $t \in [0, 1]$. If m is chosen suitable as a function of $n(\bar{X})$ then $E(\check{X}_m(t) - \bar{X}(t))^2$ is the dominating term asymptotically.

In the sequel we let c denote unspecified positive constants, which only depend on the constant K from conditions (A) as well as on $a(0, 0)$, $\sigma(0, 0)$, and $E(X(0))^4$.

8.1. *Approximation of Brownian Bridges.* Let B denote a Brownian bridge on an interval $[S, T]$ with mean zero and variance $(T - t) \cdot (t - S)/(T - S)$ at t. Let $\tilde{B}_m$ denote the piecewise linear interpolation of B at the points

$(T-S)\cdot \ell/(m+1)$ where $\ell = 1, ..., m$. Then straightforward calculations yield

$$E\left(\int_S^T (B(t)-\tilde{B}_m(t))^2\,dt\right)=\frac{(T-S)^2}{6\cdot(m+1)}. \tag{15}$$

8.2. *Proof of the Lower Bounds in Theorem* 2 *and Proposition* 2. Consider an arbitrary sequence of methods $\bar{X}_N \in \mathbb{X}_N^{**}$. Take a sequence of positive integers k_N such that

$$\lim_{N\to\infty} N^{1/2}/k_N = \lim_{N\to\infty} k_N/N = 0.$$

Since $k_N = o(N)$ we may assume that $\bar{X}_N$ uses in particular the knots

$$t_\ell = \ell/k_N, \qquad \ell = 0, ..., k_N.$$

Let $\mathfrak{A}$ denote the σ-algebra that is generated by $\Psi(X(0), W)$. Moreover, put $U_\ell = (t_\ell, \check{X}_{k_N}(t_\ell))$.

LEMMA 1.

$$\liminf_{N\to\infty} N\cdot e(\bar{X}_N)^2 \geqslant \liminf_{N\to\infty} N\cdot \sum_{\ell=0}^{k_N-1}\int_{t_\ell}^{t_{\ell+1}} E(\sigma^2(U_\ell)\cdot(W(t) - E(W(t)\mid\mathfrak{A}))^2)\,dt.$$

Proof. Theorem 4 implies

$$\int_0^1 E(X(t)-\check{X}_{k_N}(t))^2\,dt \leqslant c/k_N^2 = o(N^{-1}).$$

It is therefore sufficient to analyze $\check{X}_{k_N}-\bar{X}_N$.

Clearly

$$E(\check{X}_{k_N}(t)-\bar{X}_N(t))^2 \geqslant E(\check{X}_{k_N}(t)-E(\check{X}_{k_N}(t)\mid\mathfrak{A}))^2.$$

Let $t\in[t_\ell, t_{\ell+1}]$. Since U_ℓ is $\mathfrak{A}$-measurable we get

$$\begin{aligned}&\check{X}_{k_N}(t)-E(\check{X}_{k_N}(t)\mid\mathfrak{A})\\ &=\sigma(U_\ell)\cdot(W(t)-E(W(t)\mid\mathfrak{A}))\\ &+1/2\cdot(\sigma\cdot\sigma^{(0,1)})(U_\ell)\cdot((W(t)-W(t_\ell))^2-E((W(t)-W(t_\ell))^2\mid\mathfrak{A})).\end{aligned}$$

By boundedness of $\sigma^{(0,1)}$ and linear growth of σ,

$$\begin{aligned} E((\sigma\cdot\sigma^{(0,1)})(U_\ell)\cdot((W(t)-W(t_\ell))^2-E((W(t)-W(t_\ell))^2\mid\mathfrak{A})))^2 \\ \leqslant c\cdot E(\sigma^2(U_\ell)\cdot(W(t)-W(t_\ell))^4) \\ \leqslant c\cdot(1+E(\check{X}_{k_N}(t_\ell))^2)\cdot E(W(t)-W(t_\ell))^4 \\ \leqslant c/k_N^2, \end{aligned}$$

where the last estimate follows from Lemma 11. We conclude that

$$\begin{aligned} &\left(\int_0^1 E(\check{X}_{k_N}(t)-\bar{X}_N(t))^2\,dt\right)^{1/2} \\ &\quad\geqslant\left(\sum_{\ell=0}^{k_N-1}\int_{t_\ell}^{t_{\ell+1}} E(\sigma^2(U_\ell)\cdot(W(t)-E(W(t)\mid\mathfrak{A}))^2)\,dt\right)^{1/2}-c/k_N. \end{aligned}$$

This completes the proof, since $k_N^{-1}=o(N^{-1/2})$. ∎

According to Lemma 1 the error of every method $\bar{X}_N$ is bounded from below by an error for weighted L_2-approximation of the Brownian motion. The weight function is a stochastic process itself, and it is close to $\sigma^2(t, X(t))$. The set of observation points is determined by $\bar{X}_N$, namely,

$$D(X(0), W)=\{\psi_1(X(0)), ..., \psi_{\nu(X(0),W)}(\Psi_{\nu(X(0),W)-1}(X(0), W))\}.$$

Let

$$d_\ell(X(0), W)=\#(D(X(0), W)\cap\,]t_\ell, t_{\ell+1}[\,)+1,$$

where $\#$ denotes the cardinality of a set. Observe that $d_\ell(X(0), W)$ is $\mathfrak{A}$-measurable. Put

$$A_{k_N}(X(0), W)=\sum_{\ell=0}^{k_N-1}\frac{\sigma^2(t_\ell, X(t_\ell))}{d_\ell(X(0), W)}.$$

LEMMA 2.

$$\begin{aligned} &\liminf_{N\to\infty} N\cdot\sum_{\ell=0}^{k_N-1}\int_{t_\ell}^{t_{\ell+1}} E(\sigma^2(U_\ell)\cdot(W(t)-E(W(t)\mid\mathfrak{A}))^2)\,dt \\ &\quad\geqslant\liminf_{N\to\infty} N/(6k_N^2)\cdot E(A_{k_N}(X(0), W)). \end{aligned}$$

Proof. Clearly

$$E(\sigma^2(U_\ell)\cdot(W(t)-E(W(t)\,|\,\mathfrak{A}))^2\,|\,\mathfrak{A})=\sigma^2(U_\ell)\cdot E((W(t)-E(W(t)\,|\,\mathfrak{A}))^2\,|\,\mathfrak{A}).$$

Conditioned on $\mathfrak{A}$, the discretization $D(X(0), W)$ is fixed and the process $W(t)-E(W(t)\,|\,\mathfrak{A})$ is a Brownian bridge on each of the corresponding subintervals. Hence

$$\int_{t_\ell}^{t_{\ell+1}} E((W(t)-E(W(t)\,|\,\mathfrak{A}))^2\,|\,\mathfrak{A})\,dt\geqslant\frac{1}{6k_N^2\cdot d_\ell(X(0),W)},$$

due to (15), and consequently

$$\int_{t_\ell}^{t_{\ell+1}} E(\sigma^2(U_\ell)\cdot(W(t)-E(W(t)\,|\,\mathfrak{A}))^2)\,dt\geqslant 1/(6k_N^2)\cdot E\left(\frac{\sigma^2(U_\ell)}{d_\ell(X(0),W)}\right).$$

By (A),

$$|\sigma^2(U_\ell)-\sigma^2(t_\ell,X(t_\ell))|\leqslant c\cdot|\check{X}_{k_N}(t_\ell)-X(t_\ell)|\cdot(1+|\check{X}_{k_N}(t_\ell)|+|X(t_\ell)|).$$

Theorem 4, Lemma 11 and (10) yield

$$E\,|\sigma^2(U_\ell)-\sigma^2(t_\ell,X(t_\ell))|\leqslant c/k_N. \tag{16}$$

Since $k_N^{-2}=o(N^{-1})$ and $d_\ell(X(0),W)\geqslant 1$, the lemma follows. ∎

Now we separately analyze the three different classes $\mathbb{X}^{**}$, $\mathbb{X}^{*}$, and $\mathbb{X}$, as well as the class of methods that are based on equidistant discretizations.

LEMMA 3. *If* $\bar{X}_N=\phi_N(X(0), W(1/N), ..., W(1))$ *for every* N *then*

$$\liminf_{N\to\infty} N/k_N^2\cdot E(A_{k_N}(X(0),W))\geqslant(C^{\mathrm{equi}})^2.$$

Proof. Clearly, the numbers $d_\ell(X(0), W)$ do not depend on $X(0)$ and W, and we have

$$d_\ell\leqslant N/k_N+1.$$

Hence

$$N/k_N^2 \cdot E(A_{k_N}(X(0), W)) \geqslant N/k_N^2 \cdot E\left(\sum_{\ell=0}^{k_N-1} \frac{\sigma^2(t_\ell, X(t_\ell))}{(N+k_N)/k_N}\right)$$
$$= (N/(N+k_N)) \cdot \frac{1}{k_N} \sum_{\ell=0}^{k_N-1} E(\sigma^2(t_\ell, X(t_\ell))).$$

Since $k_N = o(N)$ the right hand side tends to $(C^{\text{equi}})^2$. ∎

LEMMA 4. *If $\bar{X}_N \in \mathbb{X}_N$ for every N then*

$$\liminf_{N\to\infty} N/k_N^2 \cdot E(A_{k_N}(X(0), W)) \geqslant C^2.$$

Proof. By definition of $\mathbb{X}_N$, the numbers $d_\ell(X(0), W)$ do not depend on $X(0)$ and W, and

$$\sum_{\ell=0}^{k_N-1} d_\ell \leqslant N.$$

The Cauchy–Schwarz inequality yields

$$N/k_N^2 \cdot E(A_{k_N}(X(0), W)) \geqslant 1/k_N^2 \cdot \sum_{\ell=0}^{k_N-1} E\left(\frac{\sigma^2(t_\ell, X(t_\ell))}{d_\ell}\right) \cdot \sum_{\ell=0}^{k_N-1} d_\ell$$
$$\geqslant 1/k_N^2 \cdot \left(\sum_{\ell=0}^{k_N-1} (E(\sigma^2(t_\ell, X(t_\ell))))^{1/2}\right)^2.$$

The right-hand side tends to C^2. ∎

LEMMA 5. *If $\bar{X}_N \in \mathbb{X}_N^*$ for every N then*

$$\liminf_{N\to\infty} N/k_N^2 \cdot E(A_{k_N}(X(0), W)) \geqslant (C^*)^2.$$

Proof. By definition of $\mathbb{X}_N^*$,

$$\sum_{\ell=0}^{k_N-1} d_\ell(X(0), W) \leqslant N.$$

The Cauchy–Schwarz inequality yields

$$N \cdot A_{k_N}(X(0), W) \geqslant \sum_{\ell=0}^{k_N-1} \frac{\sigma^2(t_\ell, X(t_\ell))}{d_\ell(X(0), W)} \cdot \sum_{\ell=0}^{k_N-1} d_\ell(X(0), W)$$
$$\geqslant \left(\sum_{\ell=0}^{k_N-1} |\sigma|(t_\ell, X(t_\ell)) \right)^2.$$

Thus

$$\liminf_{N \to \infty} N/k_N^2 \cdot E(A_{k_N}(X(0), W))$$
$$\geqslant E\left(\lim_{N \to \infty} 1/k_N \cdot \sum_{l=0}^{k_N-1} |\sigma|(t_\ell, X(t_\ell))| \right)^2 = (C^*)^2$$

by Fatou's Lemma and by continuity of X and σ. ∎

LEMMA 6. *If $\bar{X}_N \in \mathbb{X}_N^{**}$ for every N then*

$$\liminf_{N \to \infty} N/k_N^2 \cdot E(A_{k_N}(X(0), W)) \geqslant (C^{**})^2.$$

Proof. By definition of $\mathbb{X}_N^{**}$,

$$\sum_{\ell=0}^{k_N-1} E(d_\ell(X(0), W)) \leqslant N.$$

The Cauchy–Schwarz inequality yields

$$N \cdot E(A_{k_N}(X(0), W)) \geqslant \sum_{\ell=0}^{k_N-1} E\left(\frac{\sigma^2(t_\ell, X(t_\ell))}{d_\ell(X(0), W)} \right) \cdot \sum_{\ell=0}^{k_N-1} E(d_\ell(X(0), W))$$
$$\geqslant \left(\sum_{\ell=0}^{k_N-1} \left(E\left(\frac{\sigma^2(t_\ell, X(t_\ell))}{d_\ell(X(0), W)} \right) \right)^{1/2} \cdot (E(d_\ell(X(0), W)))^{1/2} \right)^2$$
$$\geqslant \left(\sum_{\ell=0}^{k_N-1} E(|\sigma|\,(t_\ell, X(t_\ell))) \right)^2.$$

Thus

$$\liminf_{N\to\infty} N/k_N^2 \cdot E(A_{k_N}(X(0), W))$$

$$\geqslant \left(\lim_{N\to\infty} 1/k_N \cdot \sum_{\ell=0}^{k_N-1} E(|\sigma|\,(t_\ell, X(t_\ell))) \right)^2 = (C^{**})^2,$$

as claimed. ∎

We combine Lemmas 1 and 2 with Lemma 3–6 to obtain the lower bounds in Theorem 2 and Proposition 2. Moreover, if N is chosen appropriately then these lower bounds yield the lower bounds in Theorem 1 and Proposition 1.

8.3. *Proof of the Upper Bounds in Theorem* 1 *and Proposition* 1. First we consider the method $\hat{X}_h^{**}$. Let

$$t_\ell = \tau_{\ell,0} = \ell/k_h$$

and consider the corresponding process $\check{X}_{k_h}$. Observe that $\check{X}_{k_h}$ and $\hat{X}_h^{**}$ coincide at the points t_ℓ. Recall that $\sigma_\ell = \sigma(t_\ell, \check{X}_{k_h}(t_\ell))$.

Asymptotically, the expected number of observations $n(\hat{X}_h^{**})$ increases like $1/h \cdot C^{**}$.

LEMMA 7.

$$\lim_{h\to 0} h \cdot n(\hat{X}_h^{**}) = \lim_{h\to 0} 1/k_h \cdot \sum_{\ell=0}^{k_h-1} E(|\sigma_\ell|) = C^{**}.$$

Proof. Note that

$$n(\hat{X}_h^{**}) \leqslant k_h + 1/(h \cdot k_h) \cdot \sum_{\ell=0}^{k_h-1} E(|\sigma_\ell|).$$

Hence $n(\hat{X}_h^{**}) < \infty$ due to (A) and Lemma 11. Moreover,

$$E\,|\sigma_\ell - \sigma(t_\ell, X(t_\ell))| \leqslant c/k_h$$

due to (A) and Theorem 4. Clearly $\lim_{h \to 0} k_h = \infty$ by (6). Use (5) to conclude that

$$\limsup_{h \to 0} h \cdot n(\hat{X}_h^{**}) \leqslant \limsup_{h \to 0} 1/k_h \cdot \sum_{\ell=0}^{k_h - 1} E(|\sigma_\ell|) \leqslant C^{**}.$$

The lower bounds are established in a similar way. ∎

Because of Lemma 7 the upper bound in Theorem 1(i) reads

$$\limsup_{h \to 0} h^{-1} \cdot e(\hat{X}_h^{**})^2 \leqslant C^{**}/6.$$

Theorem 4 and (6) imply

$$E\left(\int_0^1 (X(t) - \check{X}_{k_h}(t))^2 \, dt\right) \leqslant c/k_h^2 = o(h).$$

Hence the upper bound in Theorem 1(i) is a consequence of the following estimate.

LEMMA 8.

$$\limsup_{h \to 0} h^{-1} \cdot E\left(\int_0^1 (\check{X}_{k_h}(t) - \hat{X}_h^{**}(t))^2 \, dt\right) \leqslant C^{**}/6.$$

Proof. Fix $\ell \in \{0, ..., k_h - 1\}$ and let $\mathfrak{B}$ denote the σ-algebra that is generated by $X(0)$, $W(t_1)$, ..., $W(t_\ell)$. Define

$$d_\ell = \max(1, \lceil |\sigma_\ell|/(h \cdot k_h) \rceil)$$

and

$$V_h(t) = 1/2 \cdot (\sigma \cdot \sigma^{(0,1)})(t_\ell, \check{X}_{k_h}(t_\ell)) \cdot ((W(t) - W(t_\ell))^2 - (t - t_\ell)) \tag{17}$$

for $t \in [t_\ell, t_{\ell+1}[$. The adaptive discretization yields points

$$t_\ell = \tau_{\ell,0} < \tau_{\ell,1} < \cdots < \tau_{\ell,d_\ell - 1} < \tau_{\ell,d_\ell} = t_{\ell+1},$$

which are measurable with respect to $\mathfrak{B}$. Let $\tilde{W}_h$ denote the piecewise linear interpolation of $W - W(t_\ell)$ at these points.

Note that

$$\check{X}_{k_h}(t) - \hat{X}_h^{**}(t) = \sigma_\ell \cdot (W(t) - W(t_\ell) - \tilde{W}_h(t)) + V_h(t)$$

if $t_\ell \leqslant t \leqslant \tau_{\ell, d_\ell - 1}$, and

$$|\check{X}_{k_h}(t) - \hat{X}_h^{**}(t)| \leqslant |\sigma_\ell| \cdot |W(t) - W(t_\ell) - \tilde{W}_h(t)| + |V_h(t)| + |V_h(t_{\ell+1})|$$

if $\tau_{\ell, d_\ell - 1} < t < t_{\ell+1}$. Conditioned on $\mathfrak{B}$ the discretization is fixed and the process $W(t) - W(t_\ell) - \tilde{W}_h(t)$ is a Brownian Bridge on each of the subintervals $[\tau_{\ell, j}, \tau_{\ell, j+1}]$. Hence

$$\int_{t_\ell}^{t_{\ell+1}} E(\sigma_\ell^2 \cdot (W(t) - W(t_\ell) - \tilde{W}_h(t))^2 \mid \mathfrak{B})\, dt \leqslant (d_\ell - 1) \cdot h^2/6 + c \cdot h^2$$

$$\leqslant |\sigma_\ell| \cdot h/(6 \cdot k_h) + c \cdot h^2$$

by (15), and we obtain

$$h^{-1} \cdot \sum_{\ell=0}^{k_h - 1} \int_{t_\ell}^{t_{\ell+1}} E(\sigma_\ell^2 \cdot (W(t) - W(t_\ell) - \tilde{W}_h(t))^2)\, dt$$

$$\leqslant 1/(6 \cdot k_h) \cdot \sum_{\ell=0}^{k_h - 1} E(|\sigma_\ell|) + c \cdot h,$$

which tends to $C^{**}/6$, see Lemma 7.

It remains to observe that

$$\sup_{t \in [0,1]} E(|V_h(t)|^2) \leqslant c/k_h^2 = o(h)$$

follows from (A), Lemma 11, and (6). ∎

Now we analyze the method $\hat{X}_n^*$. Let

$$t_\ell = \tau_{\ell, 0} = \ell/k_n$$

and consider the corresponding process $\check{X}_{k_n}$. Due to Theorem 4 and (9) the upper bound in Theorem 1(ii) is a consequence of the following estimate.

LEMMA 9.

$$\limsup_{n\to\infty} n\cdot E\left(\int_0^1 (\check{X}_{k_n}(t)-\hat{X}_n^*(t))^2\,dt\right)\leqslant (C^*)^2/6.$$

Proof. We proceed as in the proof of Lemma 8. Let $t\in[t_\ell, t_{\ell+1}[$ and define V_n by the right-hand side of (17) with k_n instead of k_h. Then $E(|V_n(t)|^2)=o(n^{-1})$ uniformly in $t\in[0,1]$. Let $\mathfrak{B}$ denote the σ-algebra that is generated by $X(0)$, $W(t_1)$, ..., $W(t_\ell)$, and let $\tilde{W}_n$ denote the piecewise linear interpolation of $W-W(t_\ell)$ at the points $\tau_{\ell,0}$, ..., $\tau_{\ell+1,0}$. We use (15) to obtain

$$\int_{t_\ell}^{t_{\ell+1}} E(\sigma_\ell^2\cdot(W(t)-W(t_\ell)-\tilde{W}_n(t))^2\mid\mathfrak{B})\,dt\leqslant\frac{|\sigma_\ell|\cdot\sum_{i=0}^{k_n-1}|\sigma_i|}{6\cdot k_n^2\cdot(n-k_n)}.$$

Thus

$$n\cdot\sum_{\ell=0}^{k_n-1}\int_{t_\ell}^{t_{\ell+1}} E(\sigma_\ell^2\cdot(W(t)-W(t_\ell)-\tilde{W}_n(t))^2)\,dt\leqslant\frac{E(\sum_{i=0}^{k_n-1}|\sigma_i|)^2}{6\cdot k_n^2}\cdot\frac{n}{n-k_n}.$$

Due to (A), Theorem 4, and (8) the right-hand side tends to $(C^*)^2/6$. ∎

Next we turn to the method $\hat{X}_n$. Recall that $\hat{X}_n$ coincides with the method $\hat{X}_n^*$ except for the deterministic choice of the numbers μ_ℓ. In view of the arguments given above, for the upper bound in Theorem 1(iii) it thus suffices to prove the following estimate.

LEMMA 10.

$$\limsup_{n\to\infty} n/k_n^2\cdot\sum_{\ell=0}^{k_n-1} E(\sigma_\ell^2)/(\mu_\ell+1)\leqslant C^2.$$

Proof. Observing (16) and the definition of μ_ℓ we have

$$E(\sigma_\ell^2)/(\mu_\ell+1)\leqslant(\alpha^2(t_\ell)+c/k_n)/(\mu_\ell+1)\leqslant\alpha(t_\ell)\cdot\sum_{i=0}^{k_n-1}\alpha(t_i)/(n-k_n)+c/k_n.$$

Hence

$$n/k_n^2\cdot\sum_{\ell=0}^{k_n-1} E(\sigma_\ell^2)/(\mu_\ell+1)\leqslant\left(1/k_n\cdot\sum_{\ell=0}^{k_n-1}\alpha(t_\ell)\right)^2\cdot n/(n-k_n)+c\cdot n/k_n^2.$$

Now use (8) and (9) to complete the proof. ∎

Finally, we turn to the upper bound for the equidistant Milstein scheme $\hat{X}_n^{\text{equi}}$. Due to Theorem 4 it suffices to show that

$$\limsup_{n\to\infty} n^{1/2}\cdot\left(\int_0^1 E(\check{X}_n(t)-\hat{X}_n^{\text{equi}}(t))^2\,dt\right)^{1/2}\leqslant C^{\text{equi}}/\sqrt{6}. \tag{18}$$

Proof. Recall that $\check{X}_n$ coincides with $\hat{X}_n$ at the discretization points t_ℓ and let $\tilde{W}_n$ denote the piecewise linear interpolation of W at these points. We have

$$\begin{aligned}\check{X}_n(t)-\hat{X}_n^{\text{equi}}(t)&=\sigma(U_\ell)\cdot(W(t)-\tilde{W}_n(t))\\&\quad+1/2\cdot(\sigma\cdot\sigma^{(0,1)})(U_\ell)\cdot((W(t)-W(t_\ell))^2-n\cdot(t-t_\ell)\\&\quad\cdot(W(t_{\ell+1})-W(t_\ell))^2)\end{aligned}$$

for $t\in[t_\ell, t_{\ell+1}]$.

Note that $W(t)-\tilde{W}_n(t)$ and U_ℓ are independent if $t\geqslant t_\ell$, and

$$\begin{aligned}E((\sigma\cdot\sigma^{(0,1)})(U_\ell)\cdot((W(t)-W(t_\ell))^2-n\cdot(t-t_\ell)\\\cdot(W(t_{\ell+1})-W(t_\ell))^2))^2\leqslant c/n^2,\end{aligned}$$

similarly to the proof of Lemma 1.

Hence

$$\begin{aligned}&n^{1/2}\cdot\left(\int_0^1 E(\check{X}_n(t)-\hat{X}_n^{\text{equi}}(t))^2\,dt\right)^{1/2}\\&\leqslant n^{1/2}\cdot\left(\int_0^1 E(\sigma^2(U_\ell)\cdot(W(t)-\tilde{W}_n(t))^2)\,dt\right)^{1/2}+c/n^{1/2}\\&=n^{1/2}\cdot\left(\sum_{\ell=0}^{n-1}E(\sigma^2(U_\ell))\cdot\int_{t_\ell}^{t_{\ell+1}}E(W(t)-\tilde{W}_n(t))^2\,dt\right)^{1/2}+c/n^{1/2}\\&=1/\sqrt{6}\cdot\left(1/n\cdot\sum_{\ell=0}^{n-1}E(\sigma^2(U_\ell))\right)^{1/2}+c/n^{1/2}\\&\leqslant 1/\sqrt{6}\cdot\left(1/n\cdot\sum_{\ell=0}^{n-1}E(\sigma^2(t_\ell, X(t_\ell)))\right)^{1/2}+c/n^{1/2}\end{aligned}$$

by (15), (A), and Theorem 4. The right hand side tends to $C^{\text{equi}}/\sqrt{6}$. ∎

Obviously the upper bounds from Theorem 1 and Proposition 1 imply the upper bounds from Theorem 2 and Proposition 2.

APPENDIX

Choose fixed points

$$0 = t_0 < \cdots < t_m = 1$$

and consider the process $\check{X} = \check{X}_m$ for this discretization; see (14). Put

$$\Delta_\ell = t_{\ell+1} - t_\ell.$$

We derive a uniform upper bound for $E(X(t) - \check{X}(t))^2$ in terms of

$$\Delta_{\max} = \max_{\ell = 0, \ldots, m-1} \Delta_\ell.$$

We rely on this estimate in the analysis of numerical methods. Recall that $\check{X}$ itself cannot be used as a numerical method for the global approximation of X.

Faure (1992) presents the upper bound in the special case of constant step-size $\Delta_{\max} = 1/m$. Our proof is completely different from his approach. Moreover, it is based on weaker assumptions (see Section 3). In particular we do not need the existence of second derivatives with respect to the state variable.

As previously, c denotes unspecified positive constants, which only depend on the constant K from (A) as well as on $a(0, 0)$, $\sigma(0, 0)$, and $E(X(0))^4$.

LEMMA 11.

$$\sup_{t \in [0, 1]} E(\check{X}(t))^4 \leqslant c$$

and

$$\sup_{t \in [t_\ell, t_{\ell+1}]} E(\check{X}(t) - \check{X}(t_\ell))^4 \leqslant c \cdot \Delta_\ell^2.$$

Proof. Put $\|Y\| = (E|Y|^4)^{1/4}$ and define

$$Z(t) = \sum_{\ell=0}^{m-1} \sigma(t_\ell, \check{X}(t_\ell)) \cdot 1_{[t_\ell, t_{\ell+1}[}(t)$$

as well as

$$g(t) = \sup_{0 \leqslant s \leqslant t} \|\check{X}(s)\|.$$

Let $t \in [t_\ell, t_{\ell+1}]$. Then

$$\|\check{X}(t) - \check{X}(t_\ell)\| \leqslant c \cdot (1 + \|\check{X}(t_\ell)\|) \cdot \Delta_\ell^{1/2}, \tag{19}$$

such that $g(1) < \infty$ follows from $\|X(0)\| < \infty$. Furthermore,

$$\begin{aligned}\|\check{X}(t)\| &\leqslant c \cdot \sum_{j=0}^{\ell-1} (1 + \|\check{X}(t_j)\|) \cdot \Delta_j + c \cdot (1 + \|\check{X}(t_\ell)\|) \cdot (t - t_\ell) \\ &\quad + \left\| \int_0^t Z(s)\, dW(s) \right\| \\ &\leqslant c \cdot \left(1 + \int_0^t g(s)\, ds\right) + \left\| \int_0^t Z(s)\, dW(s) \right\|.\end{aligned}$$

Observe that

$$\|Z(t)\| \leqslant c \cdot (1 + \sup_{j=0,\ldots,\ell} \|\check{X}(t_j)\|) \leqslant c \cdot (1 + g(t)).$$

A well-known martingale moment inequality yields

$$\left\| \int_0^t Z(s)\, dW(s) \right\|^4 \leqslant 36 \cdot t \cdot \int_0^t \|Z(s)\|^4\, ds;$$

see, e.g., Karatzas and Shreve (1997, p. 163).

We conclude that

$$\|\check{X}(t)\| \leqslant c \cdot \left(1 + \left(\int_0^t g^4(s)\, ds\right)^{1/4}\right)$$

for all $t \in [0, 1]$. Thus

$$g^4(t) \leqslant c \cdot \left(1 + \int_0^t g^4(s)\, ds\right),$$

and Grownall's Lemma yields

$$g(t) \leqslant c.$$

This completes the proof of the first estimate. The second estimate is an immediate consequence of (19) and the first estimate. ▌

THEOREM 4.

$$\sup_{t \in [0,1]} E(X(t) - \check{X}(t))^2 \leqslant c \cdot \Delta_{\max}^2.$$

Proof. Put $U_\ell = (t_\ell, \check{X}(t_\ell))$. By definition, $X(t) = X(0) + A(t) + B(t)$ where

$$A(t) = \int_0^t a(s, X(s))\, ds$$

and

$$B(t) = \int_0^t \sigma(s, X(s))\, dW(s).$$

Similarly, $\check{X}(t) = X(0) + \check{A}(t) + \check{B}(t)$ where

$$\check{A}(t) = \int_0^t \sum_{\ell=0}^{m-1} a(U_\ell) \cdot 1_{]t_\ell, t_{\ell+1}]}(s)\, ds$$

and

$$\check{B}(t) = \int_0^t \sum_{\ell=0}^{m-1} \sigma(U_\ell) \cdot (1 + \sigma^{(0,1)}(U_\ell) \cdot (W(s) - W(t_\ell))) \cdot 1_{]t_\ell, t_{\ell+1}]}(s)\, dW(s).$$

For estimation of $A - \check{A}$ we define

$$Z(t) = \sum_{\ell=0}^{m-1} V_\ell(t) \cdot 1_{]t_\ell, t_{\ell+1}]}(t),$$

where

$$V_\ell(t) = a^{(0,1)}(U_\ell) \cdot (\check{X}(t) - \check{X}(t_\ell) - a(U_\ell) \cdot (t - t_\ell)).$$

Let $t \in [t_\ell, t_{\ell+1}]$. Clearly

$$\begin{aligned} a(t, X(t)) - a(U_\ell) - V_l(t) &= a(t, X(t)) - a(t_\ell, X(t)) + a(t_\ell, X(t)) - a(t_\ell, \check{X}(t)) \\ &\quad + a(t_\ell, \check{X}(t)) - a(U_\ell) - a^{(0,1)}(U_\ell) \cdot (\check{X}(t) - \check{X}(t_\ell)) \\ &\quad + a^{(0,1)}(U_\ell) \cdot a(U_\ell) \cdot (t - t_\ell). \end{aligned}$$

Lemma 11 and (10) yield

$$E(a(t, X(t)) - a(U_\ell) - V_\ell(t))^2 \leqslant c \cdot (\Delta_\ell^2 + E(X(t) - \check{X}(t))^2),$$

and therefore

$$\begin{aligned} &\int_0^t \sum_{\ell=0}^{m-1} E(a(s, X(s)) - a(U_\ell) - V_\ell(s))^2 \cdot 1_{[t_\ell, t_{\ell+1}[}(s)\, ds \\ &\qquad \leqslant c \cdot \left(\Delta_{\max}^2 + \int_0^t E(X(s) - \check{X}(s))^2\, ds \right). \end{aligned} \tag{20}$$

Note that

$$V_\ell(t) = (a^{(0,1)} \cdot \sigma)(U_\ell) \cdot \int_{t_\ell}^t (1 + \sigma^{(0,1)}(U_\ell) \cdot (W(u) - W(t_\ell)))\, dW(u).$$

If $s \in]t_\ell, t_{\ell+1}]$, too, with $s \leqslant t$, then

$$\begin{aligned} E(Z(s) \cdot Z(t)) &= E(V_\ell(s) \cdot V_\ell(t)) \\ &= \int_{t_\ell}^s E((a^{(0,1)} \cdot \sigma)(U_\ell) \cdot (1 + \sigma^{(0,1)}(U_\ell) \cdot (W(u) - W(t_\ell))))^2\, du \\ &\leqslant c \cdot E(\sigma^2(U_\ell)) \cdot (s - t_\ell) \leqslant c \cdot (s - t_\ell) \end{aligned}$$

by Lemma 11. Otherwise $E(Z(s) \cdot Z(t)) = 0$. We conclude that

$$\int_0^t \int_0^t E(Z(s) \cdot Z(u))\, ds\, du \leqslant c \cdot \Delta_{\max}^2. \tag{21}$$

Combining (20) and (21) we get

$$E(A(t)-\check{A}(t))^2 \leqslant 2\cdot\left(E\left(A(t)-\int_0^t \sum_{\ell=0}^{m-1} (a(U_\ell)-V_\ell(s))\cdot 1_{]t_\ell,\, t_{\ell+1}]}(s)\,ds\right)^2 + E\left(\int_0^t Z(s)\,ds\right)^2\right)$$

$$\leqslant c\cdot\left(\Delta_{\max}^2+\int_0^t E(X(s)-\check{X}(s))^2\,ds\right).$$

For estimation of $B-\check{B}$ we define

$$R_\ell(t)=\sigma(t,X(t))-\sigma(U_\ell)\cdot(1+\sigma^{(0,1)}(U_\ell)\cdot(W(t)-W(t_\ell))).$$

Lemma 11 and (10) yield

$$E(R_\ell(t))^2 \leqslant c\cdot(\Delta_\ell^2+E(X(t)-\check{X}(t))^2)$$

for $t\in[t_\ell, t_{\ell+1}]$. Hereby

$$E(B(t)-\check{B}(t))^2 \leqslant \int_0^t \sum_{\ell=0}^{m-1} E(R_\ell(s))^2\cdot 1_{]t_\ell,\, t_{\ell+1}]}(s)\,ds$$

$$\leqslant c\cdot\left(\Delta_{\max}^2+\int_0^t E(X(s)-\check{X}(s))^2\,ds\right). \tag{23}$$

By means of (22) and (23) we conclude that

$$E(X(t)-\check{X}(t))^2 \leqslant c\cdot\left(\Delta_{\max}^2+\int_0^t E(X(s)-\check{X}(s))^2\,ds\right)$$

for all $t\in T$. It remains to apply Gronwall's Lemma. ∎

ACKNOWLEDGMENTS

The first author (N.H.) was supported by the Deutsche Forschungsgemeinschaft (DFG:GR 876/9-2). We are grateful to the referees for their valuable and detailed remarks.

REFERENCES

1. N. Bouleau and D. Lépingle, "Numerical Methods for Stochastic Processes," Wiley, New York, 1994.
2. S. Cambanis and Y. Hu, Exact convergence rate of the Euler–Maruyama scheme, with application to sampling design, *Stochastics Stochastics Rep.* **59** (1996), 211–240.
3. J. M. C. Clark and R. J. Cameron, The maximum rate of convergence of discrete approximations, *in* "Stochastic Differential Systems" (B. Grigelionis, Ed.), Lecture Notes Control Inf. Sci., Vol. 25, pp. 162–171, Springer-Verlag, Berlin, 1980.
4. O. Faure, "Simulation du mouvement brownien et des diffusions," Thèse, ENPC, Paris, 1992.
5. J. G. Gaines and T. J. Lyons, Variable step size control in the numerical solution of stochastic differential equations, *SIAM J. Appl. Math.* **57** (1997), 1455–1484.
6. P. Hertling, Nonlinear Lebesgue and Ito integration problems of high complexity, *J. Complexity*, in press.
7. N. Hofmann, T. Müller-Gronbach, and K. Ritter, Optimal approximation of stochastic differential equations by adaptive step-size control, *Math. Comp.* **69** (2000), 1017–1034.
8. N. Hofmann, T. Müller-Gronbach, and K. Ritter, Step-size control for the uniform approximation of systems of stochastic differential equations with additive noise, *Ann. Appl. Probab.* **10**, 616–633.
9. I. Karatzas and S. E. Shreve, "Brownian Motion and Stochastic Calculus," Springer-Verlag, Berlin, 1997.
10. P. Kloeden and E. Platen, "Numerical Solution of Stochastic Differential Equations," Springer-Verlag, Berlin, 1995.
11. S. Mauthner, Step size control in the numerical solution of stochastic differential equations, *J. Comput. Appl. Math.* **100** (1998), 93–109.
12. G. N. Milstein, "Numerical Integration of Stochastic Differential Equations," Kluwer Academic, Dordrecht, 1995.
13. N. J. Newton, An efficient approximation for stochastic differential equations on the partition of symmetrical first passage times, *Stochastics Stochastics Rep.* **29** (1990), 227–258.
14. K. Ritter, "Average-Case Analysis of Numerical Problems," Lecture Notes in Mathematics, Vol. 1733, Springer-Verlag, Berlin, 2000.
15. Y. Su and S. Cambanis, Sampling designs for estimation of a random process, *Stochastic Processes Appl.* **46** (1993), 47–89.
16. D. Talay, Simulation of stochastic differential systems, *in* "Probabilistic Methods in Applied Physics" (P. Krée and W. Wedig, Eds.), Lecture Notes in Physics, Vol. 451, pp. 54–96, Springer-Verlag, Berlin, 1995.
17. J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, "Information-Based Complexity," Academic Press, New York, 1988.
18. G. W. Wasilkowski and H. Woźniakowski, On the complexity of stochastic integration, *Math. Comp.*, in press.